



Published in final edited form as:

J Am Chem Soc. 2016 February 03; 138(4): 1335–1343. doi:10.1021/jacs.5b11742.

Label-free, in-solution screening of peptide libraries for binding to protein targets using hydrogen exchange-mass spectrometry (HX-MS)

Walid S. Maaty^{1,2} and David D. Weis^{2,3,4,*}

¹Department of Nucleic Acid and Protein Structure, Agricultural Genetic Engineering Research Institute, Agricultural Research Center, Giza, Egypt

²Department of Chemistry, University of Kansas

³Department of Pharmaceutical Chemistry, University of Kansas

⁴The Ralph N. Adams Institute for Bioanalytical Chemistry, University of Kansas

Abstract

There is considerable interest in the discovery of peptide ligands that bind to protein targets. Discovery of such ligands is usually approached by screening large peptide libraries. However, the individual peptides must be tethered to a tag that preserves their individual identities (e.g. phage display or one-bead one-compound). To overcome this limitation, we have developed a method for screening libraries of label-free peptides for binding to a protein target in solution as a single batch. The screening is based on decreased amide hydrogen exchange by peptides that bind to the target. Hydrogen exchange was measured by mass spectrometry. We demonstrate the approach using a peptide library derived from the *E. coli* proteome that contained 6664 identifiable features. The library was spiked separately with a peptide spanning the calmodulin binding domain of endothelial nitric oxide synthase (eNOS, 494-513) and a peptide spanning the N-terminal twenty residues of bovine ribonuclease A (S peptide). Human calmodulin and bovine ribonuclease S (RNase S) were screened against the library. Using a novel data analysis workflow we identified the eNOS peptide as the only calmodulin binding peptide and S peptide as the only ribonuclease S binding peptide in the library.

Introduction

The discovery of novel peptide ligands against proteins targets facilitates research in disciplines ranging from basic sciences to drug and vaccine discovery. Peptides that bind to cell surface proteins can be used as cell-specific probes for imaging, either as an alternative to immunohistochemistry or in *in vivo* contexts, or for the targeted delivery of chemical agents.¹ Specific interaction surfaces between proteins can be blocked by peptides that

*Address correspondence to, 1251 Wescoe Hall Drive, University of Kansas, Lawrence KS 66045, dweis@ku.edu, +1-785-864-1377 (voice), +1-785-864-5396 (facsimile).

Supporting Information: Diagram of *E. coli* peptide library workflow (Figure S1), examples of false positive mass spectra (Figure S2), mass spectra of S peptide (Figure S3), lists of peptides and proteins identified in the *E. coli* library (Tables S1 and S2), and software parameters for screening (Tables S3 and S4).

function as inhibitors of protein-protein interactions.² Peptides also act as allosteric modulators.^{3,4} Peptides ligands can be used to define hot-spots on protein surfaces⁵ that can subsequently be explored and optimized through medicinal chemistry efforts exploiting either small molecule or peptidomimetic approaches.⁶ Screening peptide libraries against antibodies is invaluable in epitope mapping.⁷

The development of peptide libraries against a target of interest can be divided into two categories: libraries developed *in vivo* through genetic approaches and chemically synthesized libraries. The most common genetic approaches are phage display and bacterial display.⁸⁻¹⁰ Here, large libraries of random peptides ($\sim 10^{10}$) are exposed on the surfaces of phage or bacterial cells as inserts or tails within specific surface proteins. Multiple rounds of affinity selection (i.e., biopanning) are used to select amino acid sequences that have high affinity for the target. The ligands are then identified by DNA sequencing. Chemically synthesized libraries are usually prepared using combinatorial chemistry.¹ In the one-bead one-compound (OBOC) approach, peptides are synthesized combinatorially such that each individual bead has a unique sequence immobilized on its surface.¹¹ In positional scanning libraries, mixtures of combinatorially synthesized peptides are holistically screened for binding.¹² Multiple rounds of iterative screening of progressively less diverse mixtures can then produce unique peptide ligands. One advantage of chemically-synthesized libraries is that it is easy to include unnatural amino acids, those other than the twenty naturally-occurring L forms.

A number of different approaches are available to screen peptide libraries for binding to a target of interest. The approaches can either be based on direct detection of binding, indirect detection through displacement, or a functional readout such as enzymatic activity or cell viability.¹³ With small libraries, screening can be carried out one peptide at a time or with individual peptides isolated in an addressable array. For large libraries such as those generated in phage display, screening must be done in one pot. Thus the challenge becomes discovery of those peptides that bind to the target in a mixture of similar peptides that do not bind. With phage and bacterial display, multiple rounds of biopanning are used to identify the highest affinity sequences. In one-bead one-compound, the individual beads are screened for binding and mechanically sorted; the peptides that exhibit binding are subsequently identified by Edman sequencing or mass spectrometry.¹⁴

A major limitation of both peptide display and one-bead one-compound approaches is that the screened peptides must carry some type of genetic or chemical tag to facilitate identification. In the peptide display approaches, either or both the N- and C-termini are tethered; in chemically-synthesized libraries, one terminal will be tethered. Addition of these tags can interfere with binding to the target, either preventing binding or promoting artefactual binding. The current state-of-the-art does not permit the direct, one-pot screening of free peptides in solution for binding to a protein target. Our work directly addresses this limitation. Here, we demonstrate a one-pot screening approach to identify peptides from arbitrary libraries of intermediate size ($<10^4$ peptides) that bind to a specific protein target. Target binding is detected by amide hydrogen exchange mass spectrometry (HX-MS) analysis of the peptides. Another unique feature of this work is that the peptide library was generated using a proteomic approach: the peptide library was obtained by multi-enzyme

proteolysis of the *E. coli* proteome. We demonstrate proof-of-concept by selective detection of the binding of two target proteins, calmodulin and ribonuclease S, with their peptide ligands.

Experimental

Preparation of the *E. coli* peptidomic library

E. coli BL21(DE3) transformed with a pET22b plasmid described previously¹⁵ was grown as follows. A single colony was selected from an ampicillin/LB agar plate that had been streaked with the glycerol stock and incubated overnight at 37 °C. The colony was used to inoculate 10 mL of LB medium containing 100 µg mL⁻¹ ampicillin. The starter culture was grown for 10 hr at 37 °C in an orbital shaker operating at 225 rpm. The starter culture was added to 200 mL of LB medium containing 100 µg mL⁻¹ ampicillin. The culture was incubated at 37 °C in an orbital shaker operating at 225 rpm. Cells were harvested at OD₆₀₀ = 1.0. The culture was split into 40 mL portions and pelleted at 10000 × *g* for 10 min at 4 °C. After discarding the supernatant, the pellets were stored at -80 °C. Cell pellets were resuspended in 50 mM ammonium bicarbonate buffer (pH 8.0) containing 10% acetonitrile by volume, protease inhibitors (cocktail set VII, Calbiochem, San Diego, CA, USA), and 10 units of DNase-RNase mix (GE Healthcare Biosciences, Pittsburgh, PA, USA). The cells were subjected to three freeze-thaw cycles (liquid nitrogen/37 °C) then briefly probe sonicated (Microson ultrasonic cell disruptor XL2000, Misonix Inc., Farmingdale, NY, USA) on ice for 5 s using lowest power setting. The lysate volume was diluted 1:1 with 6 M urea in 50 mM ammonium bicarbonate buffer (pH 8.0). The cell lysate was then subjected to three additional freeze-thaw cycles. The crude lysate was clarified by centrifugation at 17000× *g* for 20 min at 4 °C. Proteins were precipitated from the clarified lysate with 4 volumes acetone at -20 °C. The total protein content of the acetone-precipitated fraction was estimated using the bicinchoninic acid assay (ThermoFisher Scientific, Rockford, IL, USA).

Proteins in the acetone-precipitated fraction were digested with several different proteases using a modified version of the filter-aided sample preparation (FASP) process.¹⁶⁻²⁰ Precipitated proteins were dissolved in 50 mM ammonium bicarbonate buffer (pH 8.0) containing 10% acetonitrile and 6 M urea, reduced with dithiothreitol (40 mM for 1 hr at 37 °C) and then alkylated with iodoacetamide (80 mM for 0.5 hr at 23°C in the dark). Three protein samples (350 µg each) were loaded onto separate centrifugal microconcentrators (10 kDa cut-off, Low-binding Microcon Centrifugal Filter Devices, Ultracel YM-10, Millipore, Bedford, MA, USA). The sample buffer, containing reducing and alkylating agents, was removed by centrifugation. In the following steps, the liquid sample was collected from the concentrator by centrifugation at 14,000×*g* at 20 °C. Centrifugation was applied until visual inspection of the concentrator showed that the elution step was complete (typically 15-30 min). For the digestion steps, peptides were collected by retaining the concentrator flow-through. The sequence of digestions steps, described in detail here, is also shown in Supporting Figure S1. The samples were digested overnight at 37 °C with endoproteases LysC (Promega, Madison, WI, USA), Glu-C (Protea Biosciences, Morgantown, WV, USA), and Asp-N (Promega, Madison, WI, USA) at an enzyme:protein ratio of 1:50 in 100 µL of

50 mM ammonium bicarbonate (pH 8.0) containing 10% acetonitrile and either 1 M urea (for Glu-C and Asp-N digestions) or 6 M urea (for Lys-C digestion). After recovering the peptides, an additional 100 μ L of digestion buffer was added and the samples were digested for an additional 3 hr at 37 °C. After recovering the peptides, trypsin (Promega, Madison, WI, USA), at an enzyme:protein ratio of 1:50, was then added to all three samples for overnight digestion at 37 °C in 100 μ L of 50 mM ammonium bicarbonate (pH 8.0) containing 10% acetonitrile with 1 M urea. After recovering the peptides, the Lys-C digested-sample was further digested with Glu-C (at an enzyme:protein ratio of 1:50) for 3 hr at 37 °C. After recovering the peptides, an additional 100 μ L of digestion buffer was added to all three samples and for 3 additional hours of digestion. Ultimately, all of the peptides liberated by centrifugation at each step were pooled into a single sample. The library was sub-aliquoted and then vacuum dried on an evaporative concentrator for 2 hours at 30 °C (Freezone model 7670521, Labconco, Kansas City, MO, USA).

LC-MS

All MS measurements were made with a quadrupole-time of flight mass spectrometer (Agilent 6530, Santa Clara, CA, USA) interfaced to a multi-pump LC system (Agilent 1200 series, Santa Clara, CA, USA). Liquid handling and valve switching was performed by a robotic liquid handler (LEAP HDX PAL, LEAP Technologies, Carborro, USA). LC separations consisted of online desalting and concentration on a C8 trap (Poroshell 120 EC-C8, 2.1 \times 5 mm, 2.7 μ m particles, Agilent, Santa Clara, CA, USA) with 0.1% formic acid at 200 μ L min⁻¹ followed by gradient elution at 200 μ L min⁻¹ through a C18 column (Zorbax 300SB-C18, 1.8 μ m particles, 2.1 \times 50 mm, Agilent Technologies, Santa Clara, CA, USA) using a water/acetonitrile gradient with both mobile phases containing 0.1% formic acid.

Peptides in the *E. coli* library were identified in a series of runs using data-dependent MS² with collision-induced dissociation. Initial feature extraction was performed in MassHunter Qualitative analysis (Agilent, Santa Clara, CA, USA). Mass spectral features were exported as .mgf files. Feature identification was performed using PeptideShaker (version 0.37.7, <http://compomics.github.io/projects/peptide-shaker.html>) proteomic informatics analysis software.²¹ Peptides were identified using a concatenated target-decoy *E. coli* FASTA-formatted protein sequences database from UniProt using the algorithms X! TANDEM, OMSSA, MyriMatch, MS-GF+ and MS Amanda. The decoy database sequences, a reversed version of the *E.coli* sequences, was created using SearchGUI (version 1.30.1, <http://compomics.github.io/projects/searchgui.html>). Validation at a 1% false discovery rate was obtained by searching against the decoy database. The search was performed in parallel using each of the proteases trypsin, LysC, GluC, AspN allowing for up to two missed cleavages, up to 8 modifications a precursor mass tolerance of 20 ppm and fragment mass tolerance of 0.05 u.

Screening

Human calmodulin, a gift from Prof. Trevor Creamer (University of Kentucky), was expressed and purified as described previously.^{22,23} The calmodulin binding domain of human endothelial nitric oxide synthase (eNOS, 494-513, hereafter “eNOS peptide”,

RKKTFKEVANAVKISASLMG), was obtained from Anaspec (Freemont, CA, USA) as a lyophilized powder at >95% purity.

Bovine ribonuclease S was obtained from Sigma (St. Louis, MO, USA). Ribonuclease S is a complex between the protein and a twenty-residue peptide (S peptide). The protein fraction was isolated from the peptide using the trichloroacetic acid precipitation method described by Richards.²⁴ RNase S at 1 mg mL⁻¹ in 20 mM HEPES / 100 mM NaCl / 10 mM CaCl₂ pH 7.5 was precipitated on ice with one-fifth volume of 20% (w/v) trichloroacetic acid and then incubated for one hour at room temperature. After centrifugation and removal of the supernatant, the RNase S pellet was reconstituted in H₂O, dialyzed overnight at 4 °C against the original HEPES buffer, and then lyophilized. Synthetic S peptide (KETAAAKFERQHMDSSSTAA) was obtained from GenScript (Piscataway, NJ, USA) as a lyophilized solid at >85% purity and used without further purification.

Calmodulin and RNase S (the target proteins) were separately screened for binding to their respective peptide ligands in the presence and absence of *E. coli*-derived peptide libraries. Prior to screening, we deuterated the peptides, either 2 μM eNOS peptide or S peptide alone or spiked into the *E. coli* library, in 25% D₂O [20 mM MES, 10 mM CaCl₂, 100 mM NaCl, pH 5.95] for 1 hr at 22 °C. Next, stock calmodulin and RNase S prepared from lyophilized solids, in 20 mM MES, 10 mM CaCl₂, 100 mM NaCl, 100 % D₂O pH 5.94 was added to reach 8 μM. In protein-free preparations, an equal volume of the buffer was substituted. Four mixtures were prepared: neat peptide, neat peptide and protein, peptide library, and peptide library and target protein. These mixtures were allowed to equilibrate for at least 1 hr at 1 °C prior to use. The combination of 25% D₂O from the peptide library and 100% D₂O from the target protein resulted in 40% D₂O in the mixtures. Labeling reactions were automated using a robotic liquid handler (LEAP HDX PAL, Carborro, NC, USA). 5 μL aliquots of the peptide stocks, held at 1 °C, were dispensed into individual autosampler vials. D → H exchange at 4 °C for 60 s was initiated by addition of 45 μL of H₂O buffer. 40 μL of the labeling mixture was quenched by mixing with an equal volume of quench buffer (200 mM glycine pH 2.5 at 1°C). After a 15 s hold at 1 °C, 70 μL of the quenched reaction mixture was loaded into the sample loop of the LC system. The temperature of the mobile phase, columns, and tubing was maintained at 1 °C using the refrigerated column compartment of the HDX PAL system. Screening reactions were run in triplicate. To minimize the complexity without complete loss of deuterium label, the labeling reactions were run using each of three different short gradients as described in further detail in the Results section resulting in a total of 36 labeling reactions (4 conditions × 3 gradients × 3 replicates) for each target protein screening.

Data analysis

Initial MS feature extraction was carried out using the Find by Molecular Feature algorithm in MassHunter Qualitative Analysis (version B.06, Agilent, Santa Clara, CA, USA). Deuterated peptide features have isotopic distribution patterns that are very different than peptides with a natural isotopic distribution. Initially, trial and error was used to adjust the algorithm parameters to reliably extract the spectra of deuterated eNOS peptide (see Table S3 in the Supporting Information). Extracted features lists, as .cef files, were further

analyzed using MassProfiler (version B.03.00, Agilent Santa Clara, CA, USA). MassProfiler is typically used to identify biomarkers in large data sets that are unique to a particular experiment. In the present case, a feature that appears at the same mass and retention time in the presence and absence of target protein was taken to be a non-binder. As with feature extraction, trial and error was required to optimize MassProfiler parameter settings to reliably return data for the eNOS peptide. (See Table S4 in the Supporting Information for the optimized parameter settings.) The output of MassProfiler was a list of features that appeared to be unique, i.e., they appeared to be present only in either the presence or absence of the target protein. This list of features was further analyzed to identify potential target binders as described in the Results section. This data analysis workflow was used without any further optimization for the screening of RNase S.

Results

Peptide library production

To test the feasibility of peptide library screening, we first developed a peptide library from *E. coli* lysate by combining sequential and parallel treatments with Lys-C, Glu-C, Asp-N, and trypsin using a filter-aided sample preparation (FASP) protocol as described in the Experimental section. In this context, *sequential* means that after treatment with a protease (e.g., Glu-C), the released peptides, as filter flow-through, were collected. The filter retentate was then treated with an additional protease (e.g., trypsin). *Parallel* means that different sequential digestions of the same lysate were carried out on separate filter units.

We used parallel FASP-based digestions starting with Asp-N, Glu-C, and Lys-C as shown in Figure S1. Trypsin was then added as a second protease to all of the filter units. In the Lys-C/trypsin digestions, Glu-C was used as a third, complimentary, protease. Peptides from all three digestions were pooled into a single peptide library containing, based on LC-MS analysis, 6200 putative peptides. The pooled library was analyzed several times by LC-MS². While our LC-MS system was not optimized for proteomic work (see Experimental) our objective here was simply to gain a general overview of the proteome coverage that could be obtained. The results from this analysis are shown in Table 1: a total of 45419 extracted MS features produced 40241 MS² spectra. Analysis using PeptideShaker produced 5646 validated MS² spectra resulting in 1541 assigned peptides from 347 different proteins in the library. The top twenty gene ontologies²⁵ and KEGG pathways²⁶ are shown in Figure 1. A complete list of the identified peptides and proteins from which they were derived are presented in Tables S1 and S2.

The protein-peptide model systems

To test the feasibility of screening for a specific protein-peptide binding within the peptidomic library, we first used the binding between calcium-activated calmodulin and a peptide spanning the calmodulin binding domain of endothelial nitric oxide synthase (eNOS, 494-513) with a dissociation constant (K_d) of 2.9×10^{-9} .²⁷ Like many other peptide ligands of calmodulin,²⁸ eNOS peptide folds into an α -helix upon binding to calcium-loaded calmodulin.^{29,30} In this context, calmodulin represents our protein target of interest. The objective of the screening was to identify peptide ligands that bound to calmodulin. eNOS

peptide was spiked into the library at an average ion abundance (11339 units) near median peptide abundance (6986 units). We used a large excess of calmodulin to drive the binding equilibrium to favor the bound state of the peptide and to ensure that the target was present in a sufficient excess that it could not be saturated by all potential peptide ligands. To demonstrate the generality of this approach and evaluate the detectable affinity range, we screened a second target protein, ribonuclease S (RNase S) using its S peptide ligand. The affinity between S peptide and RNase S, $K_d = 1 \times 10^{-7}$,³¹ is two orders of magnitude weaker than the calmodulin-eNOS peptide complex.

Gradient optimization

To minimize loss of deuterium label during the LC step, hydrogen exchange mass spectrometry workflows typically use short gradients of 15-30 minutes with only 5-10 minutes devoted to the peptide separation. The 6200 peptides in the library exceeded the practical capacity of the LC-MS system to confidently discriminate isolated isotopic clusters in the mass spectra of each peptide within the time constraints imposed by the hydrogen exchange labeling. To overcome this problem, we developed three distinct gradient profiles (i.e., early, middle, and late) to disperse separate fractions of the library into the shallow, working part of the gradient (see Table 2). The early gradient used 2-19% B over ten minutes, the middle gradient used 12-24% B over twelve minutes, and the late gradient used 26-95% B over ten minutes. The gradients and representative base peak chromatograms are shown in Figure 2. The use of three gradients allowed us to split the library into three nearly-equal size hydrophobicity fractions as summarized in Table 3. Thus the total number of detectable library peptides in each HX-MS run was decreased from 6200 to between 1431 and 2759. The sum total of features identified in the three gradients, 6664, exceeds the 6200 identified in a single, long gradient because some of the peptides probably overlap between the gradients. Although this pool of peptide features is roughly ten-fold larger than routine HX-MS experiments, we have found that useful data can still be obtained from libraries of this size.

The *E. coli*-derived peptide library is diverse with respect to mass, abundance, and hydrophobicity, as shown in Figure 3. The peptide library had a median mass of 1003 Da with a range of 300-4000 (Figure 3a). The abundance of detectable features spanned two orders of magnitude (Figure 3b). Using the optimized three-gradient separation, the features were well-dispersed across the useful retention time window of two to twenty minutes (Figure 3c). This indicates the richness and complexity of the peptide library spanning hydrophilic to hydrophobic peptides.

Optimization of labeling conditions for screening for peptide binding

Conventional HX-MS experiments typically make use of forward exchange where a protiated sample (i.e., primarily ^1H) is exposed to a large excess of D_2O to promote $\text{H} \rightarrow \text{D}$ exchange. Due to the stochastic nature of the hydrogen exchange process, forward exchange leads to an unnatural isotopic distribution that is the convolution of the natural isotopic distribution with a binomial distribution arising from the hydrogen exchange process.³² Conventional MS software is not well-suited for detection and analysis of the binomial distribution.³³ Since our goal is to use hydrogen exchange to identify peptides that bind to a

protein target, we merely need to distinguish differential protection between the bound and unbound states of the peptide. To accomplish this, we have developed a D→H exchange workflow. We started with a pre-deuterated library of peptides prepared using 40% D₂O/60% H₂O buffer. The deuterated peptides were then diluted 1:10 into H₂O in the presence and absence of the target protein. The workflow exploits rapid back-exchange by unbound peptides that causes them to undergo nearly complete back-exchange to the protiated form. Reversion to a near-natural isotopic distribution results in mass spectra with prominent monoisotopic peaks and isotopic distributions that closely resembles the natural distributions. Peptide spectra with a near-natural isotopic distribution and a well-defined monoisotopic mass peak are readily detected using mass spectrometry screening software. Since most peptides in the library do not bind to the target protein, most peptides have identical mass and nearly identical retention time in the presence and absence of the target protein. Mass spectral features that can be matched based on mass and retention time between the two conditions (i.e., presence and absence of the target protein) can be discarded as non-binders. Thus, our data analysis is reduced from examining all MS features to only identifying those few features that *appear* to be present in only one of the two states. To the screening software, the peptides appear to be unique in the two conditions because their monoisotopic masses differ by a few Da. The search for binders is reduced to cross-validation of features that were classified as unique in either of the two conditions.

Detection of binding

Figure 4 illustrates the screening approach with the eNOS peptide. During 60 seconds of D→H exchange, eNOS bound to calmodulin retains a substantial amount of deuterium (Figure 4a) while free eNOS peptide undergoes nearly complete back-exchange (Figure 4b). Even in the highly complex peptide library, the spectra of eNOS were clean and well-resolved (Figures 4cd). Based on its mass spectrum, eNOS exhibited similar protection from exchange when it bound to calmodulin in presence of the peptide library (Figure 4c) and free eNOS peptide exhibited the same level of back-exchange (Figure 4d). In contrast, other peptides exhibited identical back-exchange in both the presence and absence of calmodulin. Figures 4e and 4f show an example that illustrates the behavior of the majority of the features in the library during screening: the features had the same mass in both conditions indicating no binding to calmodulin. Since this peptide had the same mass and retention time under both conditions, it was removed from the candidate binder list. The goal of the screening workflow was to identify features that exhibited positive mass shifts in the presence of calmodulin. The entire list of peptide features was filtered to identify apparently unique peptides in the presence vs. the absence of the binding protein.

Detection of peptides that bound to calmodulin required identification of MS features with similar, but not identical masses, with nearly the same retention time. As a first pass, Mass Profiler was used to identify features that appeared to be unique in either condition. Here, unique means that a feature, detected in at least two of the three runs in one condition, could not be matched in at least two of three runs in the other condition. Table 3 summarizes the results from this analysis, applied to the screening of a library of *E. coli* peptides for binding to calmodulin. A total of 6664 features were detected in at least two of the three runs for one of the two conditions. Of these features, 406 appeared to be unique in the presence of

calmodulin while 695 appeared to be unique to the screening of the calmodulin-free library. These 1101 features were classified as potential calmodulin binding peptides. Under D \rightarrow H labeling conditions, a true binder will be more protected against exchange when it is bound to the target protein. Thus, the mass signature of a binding peptide is that it will have a higher mass in the presence of the target than a feature with nearly the same retention time in the absence of the target. To identify potential binders, the features were subjected to pair-wise comparisons (the presence vs. the absence of calmodulin) to identify potential binders based on retention time and mass proximity according to the following rules

$$|t_{\text{target}} - t_{\text{free}}| \leq 0.4 \text{ min} \quad (1)$$

$$0\text{Da} < m_{\text{target}} - m_{\text{free}} \leq 10\text{Da} \quad (2)$$

where t and m denote the retention time and mass, respectively, and the subscripts indicate the presence or absence of the target protein (here, calmodulin). 117 peptide pairs that satisfied both criteria were flagged for inspection. The mass spectra the features in both states were examined to determine if they had isotopic distributions like those shown in Figure 4b and 4d. Except for eNOS peptide, all of the flagged peptides were false positives that generally arose from failures in peak picking by the Find by Molecular Feature algorithm caused by misassignment of the monoisotopic peak, from chemical noise in the mass spectrum, or from features that could not be matched to a similar feature in the other condition. A few examples of false positive results are shown in Figure S2 of the supporting information. Only the eNOS peptide (Figure 4cd) exhibited an isotopic distribution and mass shift that would be expected for a peptide undergoing differential hydrogen exchange in the presence/absence of calmodulin.

As summarized in Table 3, this workflow led to the identification of only a single peptide ligand, eNOS, against the target protein calmodulin within a complex sample library composed of ~6000 detectable features. The true makeup of the peptide library, derived from proteolysis of *E. coli* lysate, is certainly more complex than these ~6000 features analyzed. Limits are imposed by chemical noise, the dynamic range of the mass spectrometer, and the combined resolving power of the LC and MS dimensions.

A limited screening of the *E. coli* library spiked with S peptide against RNase S was used to assess the effects of weaker binding and to determine if the screening process was directly transferable to another protein-ligand system. In this case, the elution of the S peptide in the early gradient was known *a priori*, so only the early gradient was evaluated. As with the eNOS peptide, S peptide also exhibited a higher level of deuteration in the presence of ribonuclease S (see Figure S3 in the Supporting Information), but the mass shift was smaller. The smaller mass shift is consistent with the lower affinity, $K_d \approx 1 \times 10^{-7}$ of RNase S for its ligand.³¹ Screening and data analysis were conducted without any additional optimization. The results of the RNase S screening are presented in Table 4. As with calmodulin

screening, only a single hit, S peptide, was ultimately confirmed. In RNase S screening, although the same *E. coli* proteome and proteolysis process was used for library production, fewer total peptides were detected in the early gradient (874 vs. 2474). The apparent loss of peptides may arise from differences in total lysate digested, differences in enzyme efficiency, or variability in sample loading. This lack of reproducibility result highlights a challenge in preparation of peptide libraries directly from cell lysates.

Discussion

Using ligand screening strategies such as phage display and one-bead one-compound enables the rapid screening of libraries of random peptides as large as 10^{10} . However, there are significant drawbacks to these approaches that arise from the need to tether the peptides. In phage display, for example, the peptides are fused to the phage coat protein. Constraining one or usually both termini and placing the entire peptide close to the coat surface can lead to the presentation of non-native conformations, prevent exposure of a viable binding interface, or prevent the peptide from adopting a binding-competent conformation. Thus the screening of free peptides in solution is attractive, because the solution conformation of the peptide ligand is preserved. In this work, we have demonstrated that free peptides in small libraries can be directly screened for binding to a target of interest.

Using current practices, the practical limit in HX-MS analysis seems to be pools of a few hundred peptides³⁴ as reported in challenging HX-MS experiments such as large protein complexes³⁵ or monoclonal antibody studies.³⁶ Here, hydrogen exchange analysis of 6664 peptide features is probably the largest number of peptide features analyzed in a single experimental effort. It represents, nevertheless, perhaps 2-3% of the number of peptides expected based on in silico digests of the *E. coli* proteome. Though the potential of using a proteome as a rational chemical space for peptide ligands is highly attractive, it may be difficult to use the approach outlined here to achieve the peak capacity required to fully characterize such a library. *E. coli* lysate was used here because it was a relatively simple approach for the production of a large, complex peptide library. A eukaryotic protein-peptide interaction was deliberately added to minimize the probability that a ligand would already be present in the library. It is likely that the 6664 peptides screened in the present work represent high-abundance proteins. Indeed, as shown in Figure 1, the majority of the peptides identified in our library are associated with transcription and translation.

While proteome-scale resolution is unlikely to be achievable, we do anticipate that further improvements would be able to increase capacity by factor of ten to one hundred. This would enable the screening of libraries on the order of 10^6 peptides. Increasing mass spectrometer resolution from 10^4 to 10^5 by switching from time-of-flight to orbitrap would result in a significant increase in peak capacity. Changing from narrow bore (2.1 mm diameter) LC columns operating at $200 \mu\text{L min}^{-1}$ flow rate down to capillary dimensions would increase the capacity by lowering the limit of detection. Use of capillary electrophoresis might further enhance the peak capacity.³⁷ This screening method may also be useful with libraries produced using targeted proteomic approaches such as subcellular fractionation or pull-downs of protein complexes. Finally, libraries of proteome-derived peptides could be pre-fractionated by cation exchange with each cation exchange fraction

screened separately. Furthermore, if the library was synthetic, rather than proteome-derived, the dynamic range of the sample would no longer impose such an obstacle to detection and a decrease in chemical noise from sample preparation could result in further improvement.

Ultimately, the limit of detection for binding will depend on whether there is a discernable mass difference in the presence/absence of the target. Both peptide size and affinity for the target will define this limit of detection. First, larger peptides have more amide hydrogens. Thus larger peptides will have larger mass shifts than smaller peptides. In conventional hydrogen exchange experiments, mass differences of ~0.5 Da or more are typically reported as significant,³⁸ in other words, detectable. Here however, the data analysis relies on loss of the monoisotopic peak rather than mass increase, per se. Still, the overall mass increase determines whether the monoisotopic peak is rendered undetectable. The extent of protection from hydrogen exchange will also depend on the interplay between the kinetics of peptide binding and release with the kinetics of the hydrogen exchange process. Peptide ligand, when transiently unbound, will undergo much faster hydrogen exchange, but this effect will only be important if the lifetime of the unbound peptide is comparable to lifetime of amide hydrogen exchange (approximately 0.1-10 s at pH 7).³⁹ The lifetime of unbound ligand, in turn depends on the concentration of the target protein. A large excess of target protein was used in this work to saturate the bound state of the ligands in order to minimize rapid D→H exchange by the transiently unbound peptide ligand.

In this work, our data analysis relied on the loss of monoisotopic peak to classify peptides as potential binders. Although a number of different software suites are now available for the analysis of hydrogen exchange mass spectrometry data,⁴⁰ these packages all require detailed knowledge of the peptides of interest. In particular, the amino acid sequence and retention time window for each peptide must be known in advance. Because we are screening large libraries of unknown peptides, we selected a binary classifier, a shift in the apparent monoisotopic peak, that was readily applied using off-the-shelf software. The rate of false positives, and the need for manual review, could be substantially decreased by implementation of a deuterated isotopic profile analysis of candidates that pass the initial screening.

Beyond the screening described here that was used to identify a known protein-peptide interaction, further developments in the workflow would be required to detect and identify novel peptide ligands. In particular, tandem mass spectrometry would need to be incorporated into the workflow. Since the fully back-exchanged control state is available, there would be no issues with deuterium scrambling that can be encountered during precursor fragmentation. Absence of deuteration would also facilitate peptide identification by database searching. Finally, hits identified by this technique require confirmation: interactions between peptide ligands and targets would need to be confirmed by follow-up assays such as fluorescence polarization or surface plasmon resonance.

One particular advantage of the approach described here is that screening is a one-pot process involving free peptides in solution. Thus it is not necessary to isolate the individual library members as is done in well plate assays or one-bead-one-compound screening. Furthermore, the analysis time is not a strong function of the library size as long as the peak

capacity is not exceeded. Based on this work, it would require no additional time to screen a library of 20000 instead of 6000 peptides

We anticipate several different areas where the advantages of this screening approach may be valuable. The first area is the discovery of potential leads to inhibit protein-protein interactions. Here two alternative modes are possible. In the direct mode, the discovered ligand itself is used in hit-to-lead to discover protein-protein interaction inhibitors. In the second mode, the ligand could be used as the displaceable ligand in high throughput competition assays. Beyond screening, the hydrogen exchange labeling and data analysis workflow we have described may prove useful in conventional comparative HX-MS experiments where there are only a small number of differences between two protein states. The technique might also be useful for development of designed antibodies: the screening is orthogonal to an approach in which the Protein Data Bank was mined to identify potential paratopes against β -sheet regions of disordered proteins.⁴¹

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Prof. Trevor Creamer of the University of Kentucky for the donation of calmodulin. An equipment loan from Agilent Technologies is gratefully acknowledged. Mohammed Al-Naqshabandi and Tyler Hageman provided critical readings of the manuscript. This project was supported by grant P30GM103495 from the National Institute of General Medical Sciences at the National Institutes of Health.

References

1. Gray BP, Brown KC. Chem Rev. 2014; 114:1020. [PubMed: 24299061]
2. Arkin MR, Whitty A. Curr Opin Chem Biol. 2009; 13:284. [PubMed: 19553156]
3. Goeke D, Kaspar D, Stoeckle C, Grubmüller S, Berens C, Klotzsche M, Hillen W. J Mol Biol. 2012; 416:33. [PubMed: 22178480]
4. Landgraf KE, Steffek M, Quan C, Tom J, Yu C, Santell L, Maun HR, Eigenbrot C, Lazarus RA. Nat Chem Biol. 2014; 10:567.
5. Arkin MR, Wells JA. Nat Rev Drug Discov. 2004; 3:301. [PubMed: 15060526]
6. Wells JA, McClendon CL. Nature. 2007; 450:1001. [PubMed: 18075579]
7. Liu R, Enstrom AM, Lam KS. Experimental Hematology. 2003; 31:11. [PubMed: 12543103]
8. Smith GP, Petrenko VA. Chem Rev. 1997; 97:391. [PubMed: 11848876]
9. Daugherty PS. Curr Opin Struct Biol. 2007; 17:474. [PubMed: 17728126]
10. Bratkovi T. Cellular and Molecular Life Sciences. 2010; 67:749. [PubMed: 20196239]
11. Lam KS, Salmon SE, Hersh EM, Hruby VJ, Kazmierski WM, Knapp RJ. Nature. 1991; 354:82. [PubMed: 1944576]
12. Houghten RA, Pinilla C, Blondelle SE, Appel JR, Dooley CT, Cuervo JH. Nature. 1991; 354:84. [PubMed: 1719428]
13. Lam, KS., Lehman, AL., Song, A., Doan, N., Enstrom, AM., Maxwell, J., Liu, R. Methods in enzymology. Vol. 369. Academic Press; 2003. p. 298
14. Aina OH, Liu R, Sutcliffe JL, Marik J, Pan CX, Lam KS. Molecular Pharmaceutics. 2007; 4:631. [PubMed: 17880166]
15. Keppel TR, Howard BA, Weis DD. Biochemistry. 2011; 50:8722. [PubMed: 21894929]
16. Wisniewski JR, Zougman A, Nagaraj N, Mann M. Nat Meth. 2009; 6:359.

17. Sharma R, Dill BD, Chourey K, Shah M, Ver Berkmoes NC, Hettich RL. *J Proteome Res.* 2012; 11:6008. [PubMed: 23126408]
18. Wisniewski JR, Mann M. *Anal Chem.* 2012; 84:2631. [PubMed: 22324799]
19. Erde J, Loo RR, Loo JA. *J Proteome Res.* 2014; 13:1885. [PubMed: 24552128]
20. Giannone RJ, Wurch LL, Podar M, Hettich RL. *Anal Chem.* 2015; 87:7720. [PubMed: 26110888]
21. Vaudel M, Burkhardt JM, Zahedi RP, Oveland E, Berven FS, Sickmann A, Martens L, Barsnes H. *Nat Biotech.* 2015; 33:22.
22. Charbonneau, H., Hice, R., Hart, RC., Cormier, MJ. *Methods in enzymology.* Vol. 102. Academic Press; 1983. p. 17
23. Rumi-Masante J, Rusinga FI, Lester TE, Dunlap TB, Williams TD, Dunker AK, Weis DD, Creamer TP. *J Mol Biol.* 2012; 415:307. [PubMed: 22100452]
24. Richards FM, Vithayathil PJ. *J Biol Chem.* 1959; 234:1459. [PubMed: 13654398]
25. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. *Nature genetics.* 2000; 25:25. [PubMed: 10802651]
26. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. *Nucleic Acids Res.* 2012; 40:D109. [PubMed: 22080510]
27. Matsubara M, Hayashi N, Titani K, Taniguchi H. *J Biol Chem.* 1997; 272:23050. [PubMed: 9287303]
28. Crivici A, Ikura M. *Ann Rev Biophys Biomolec Struct.* 1995; 24:85.
29. Aoyagi M, Arvai AS, Tainer JA, Getzoff ED. *EMBO J.* 2003; 22:766. [PubMed: 12574113]
30. Piazza M, Taiakina V, Guillemette SR, Guillemette JG, Dieckmann T. *Biochemistry.* 2014; 53:1241. [PubMed: 24495081]
31. Connelly PR, Varadarajan R, Sturtevant JM, Richards FM. *Biochemistry.* 1990; 29:6108. [PubMed: 2383573]
32. Arrington CB, Teesch LM, Robertson AD. *J Mol Biol.* 1999; 285:1265. [PubMed: 9887275]
33. Sarpe, V., Schriemer, DC. *Hydrogen exchange mass spectrometry: Theory, techniques, and applications.* Weis, DD., editor. Wiley; in press
34. Engen JR, Wales TE. *Annu Rev Anal Chem.* 2015; 8:127.
35. Houde D, Demarest Stephen J. *Structure.* 2011; 19:890. [PubMed: 21645859]
36. Arora J, Hickey JM, Majumdar R, Esfandiary R, Bishop SM, Samra HS, Middaugh CR, Weis DD, Volkin DB. *mAbs.* 2015; 7:525. [PubMed: 25875351]
37. Black WA, Stocks BB, Mellors JS, Engen JR, Ramsey JM. *Anal Chem.* 2015; 87:6280. [PubMed: 25992468]
38. Houde D, Berkowitz SA, Engen JR. *J Pharm Sci.* 2011; 100:2071. [PubMed: 21491437]
39. Bai Y, Milne JS, Mayne L, Englander SW. *Proteins.* 1993; 17:75. [PubMed: 8234246]
40. Iacob R, Engen J. *J Amer Soc Mass Spectrom.* 2012; 23:1. [PubMed: 22076632]
41. Sormanni P, Aprile FA, Vendruscolo M. *Proc Natl Acad Sci USA.* 2015; 112:9902. [PubMed: 26216991]

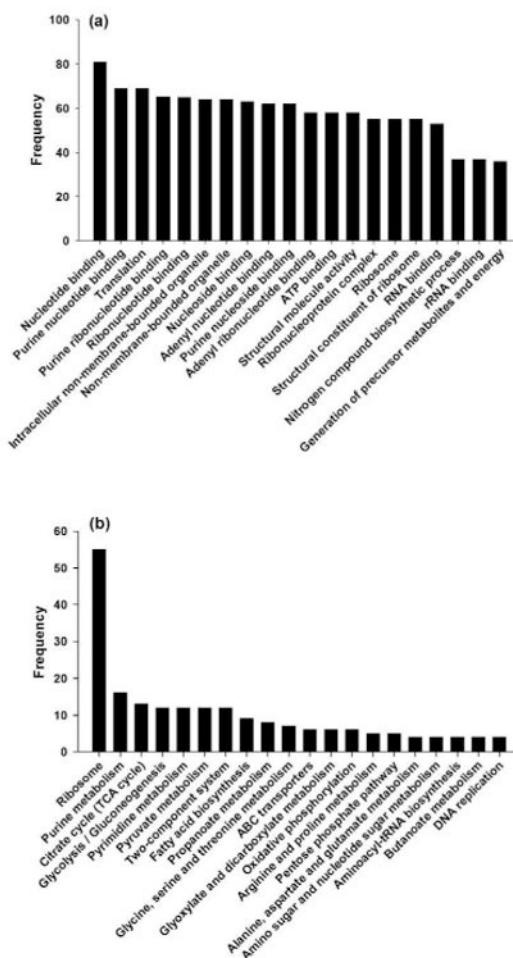


Figure 1. (a) The top twenty gene ontologies (GO terms)²⁵ and (b) top twenty pathways (KEGG annotation)²⁶ of the *E. coli* proteins identified in the peptide library.

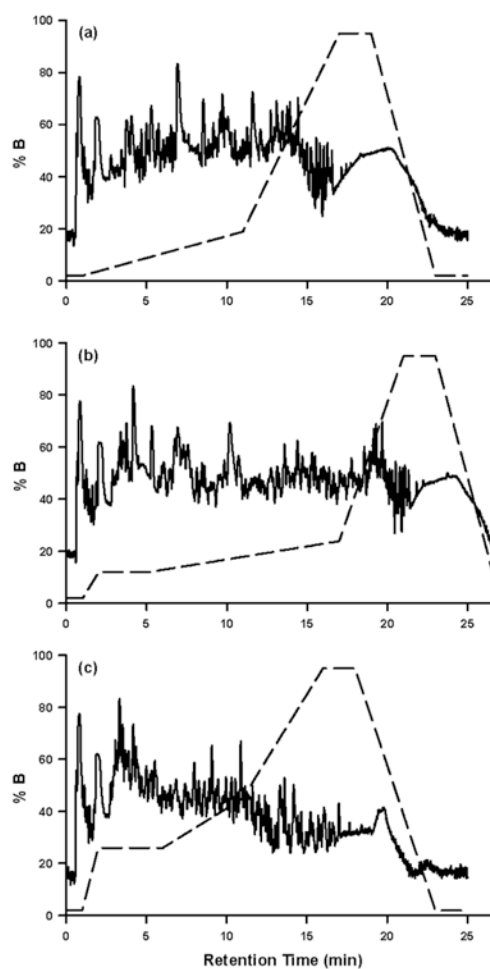


Figure 2.

Base-peak chromatograms (solid lines) and their corresponding gradients (dashed lines) for separation of the peptide library into three separate fractions, (a) early, (b) middle, (c) and late. For the base-peak chromatograms, the vertical axis is base ten logarithmic. See Table 1 for the gradient time tables and Table 2 for the number of peptides observed in each separation.

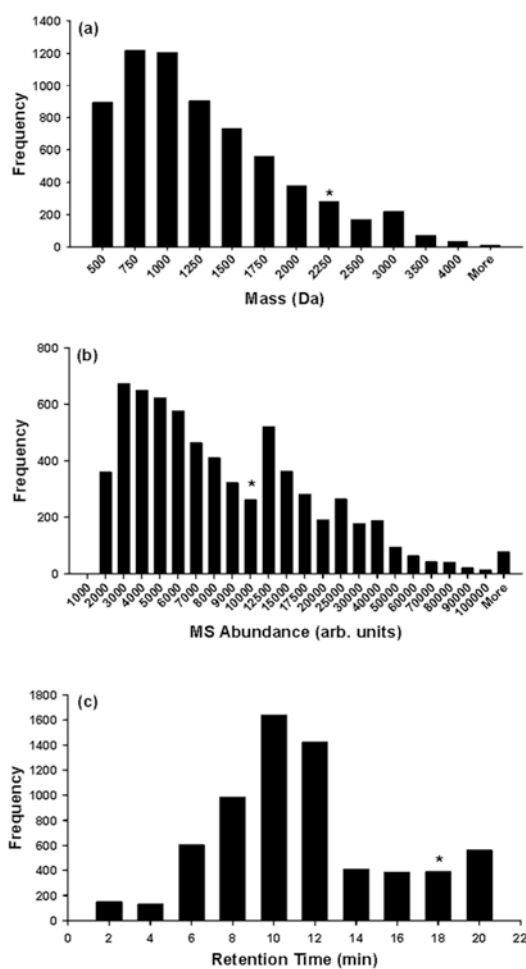


Figure 3.

Characteristics of the *E. coli* library peptides. The asterisks (*) denote the characteristics of eNOS peptide. Distributions of their (a) masses, (b) MS abundances, and (c) LC retention times. Only peptides found in the shallow portions of the gradients (see Table 2) are presented. The numbers on the horizontal axes the upper limit of the bins.

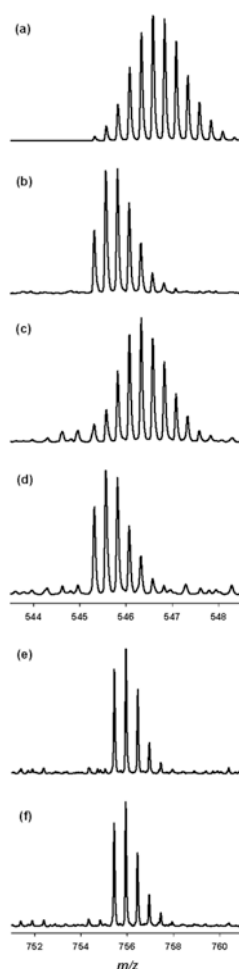


Figure 4.

Representative hydrogen exchange mass spectra showing mass shifts observed during D \rightarrow H exchange. (a)-(d) show the spectrum of the eNOS peptide, (e) and (f) are of *E. coli* D-ribose-binding periplasmic protein (P02925) peptide $^{165}\text{LAATIAQLPDQIGAK}^{179}$. In the presence of calcium-loaded calmodulin, eNOS peptide experiences less hydrogen exchange (a) than it does when calmodulin is absent (b). The hydrogen exchange characteristics of the eNOS peptide are similar when the hydrogen exchange screening is applied in the presence of the peptide library (c and d). Other peptides in the library show no difference in hydrogen exchange in the presence/absence of calmodulin (e and f).

Table 1

Analysis of the *E. coli* peptide library by LC-MS² with collision induced dissociation.

MS ² spectra acquired	40241
Validated MS ² spectra	5646
Peptides identified	1541
Proteins identified	347
GO terms	169
KEGG pathways	32

Abbreviations: GO (gene ontology); KEGG (Kyoto Encyclopedia of Genes and Genomes).

LC-MS gradients used for hydrogen exchange screening of the peptide library as shown in Figure 2. The shallow portion of the gradient is denoted by numbers in bold. The B mobile phase was 90% acetonitrile/10% water/01.% formic acid by volume. The A mobile phase (not listed) was 0.1% formic acid.

Table 2

Early Gradient						
Time (min)	0	1	11	17	19	23
%B	2	2	19	95	95	2
Middle Gradient						
Time (min)	0	1	2	5	17	21
%B	2	2	12	12	24	95
						2
Late Gradient						
Time (min)	0	1	2	6	11	16
%B	2	2	26	26	40	95
						2

Table 3

Results from screening the eNOS peptide-spiked *E. coli* peptide library for binding to calmodulin.

	Early gradient	Middle Gradient	Late Gradient	Total
Identified features	2474	2759	1431	6664
Unique (– calmodulin)	174	306	215	695
Unique (+ calmodulin)	186	125	95	406
Identified hits ¹	28	43	46	117
Confirmed hits ²	0	1	0	1

¹Hits were identified on the basis of altered hydrogen exchange using the criteria in equations (1) and (2).

²Hits were confirmed by inspection of the extracted spectra as illustrated in Figures 4 and S2.

Table 4

Results from screening the S peptide-spiked *E. coli* peptide library for binding to ribonuclease S.

	Early gradient
Identified features	874
Unique (– ribonuclease S)	51
Unique (+ ribonuclease S)	33
Identified hits ¹	4
Confirmed hits ²	1

¹Hits were identified on the basis of altered hydrogen exchange using the criteria in equations (1) and (2).

²Hits were confirmed by inspection of the extracted spectra as illustrated in Figure S3.